

The Extraction of Nonverbal Behaviors: Using Video Images and Speech-Signal Analysis in Dyadic Conversation

Ken Fujiwara · Ikuo Daibo

Published online: 29 April 2014
© Springer Science+Business Media New York 2014

Abstract This study examined the validity and possible utility of a new procedure for the extraction of nonverbal behaviors from dyadic conversation. Three methods were used to extract nonverbal behaviors (i.e., hand gestures, adaptors, and utterances). A novel automated method employing video images and speech-signal analysis software programs was compared to the more traditional coding and behavioral rating methods. The automated and coding methods provided an objective count of how many times a target behavior occurred, while behavioral ratings were based on more subjective impressions. Although there was no difference between the automated and coding methods for hand gestures, the coding method using an event recorder yielded marginally significantly more instances of adaptors and utterances as compared to the software programs. Measures of each nonverbal behavior were positively correlated across the different methods. In addition, interpersonal impressions of each speaker were rated by both observers and conversational partners. Although R^2 was lower than for the coding/behavioral rating methods, nonverbal behaviors extracted using the software programs significantly predicted familiarity and activeness ratings from both observer and partner points of view. These results support the validity and possible utility of the software-based automated extraction procedure.

Keywords Nonverbal behavior · Automated tool · Event recorder · Behavioral rating · Social interaction

K. Fujiwara (✉)
Department of Social Psychology, Graduate School of Human Sciences, Osaka University,
1-2 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: k.fujiwara1985@gmail.com

I. Daibo
Tokyo Future University, Adachi, Japan

Introduction

Nonverbal behavior is a crucial aspect of interpersonal communication. Such behaviors are involved in the expression of intimacy (Daibo 1990; Floyd 2006), impression formation (Riggio and Friedman 1986), rapport (Bernieri et al. 1996), and the expression of dominance and power in relationships (Dunber and Burgoon 2005). Studies examining nonverbal behaviors generally first extract and measure features of such behaviors and then examine the relationship between these measures and other variables of interest. It may not be an overstatement to say that these studies depend on the reliable extraction of nonverbal behavior, with coders playing a key role. These individuals carefully observe whether the target behavior occurred and code instances into analyzable data.

Unfortunately, such studies are costly and require a great deal of time and effort to conduct. Coders require initial training to be able to identify and extract specific behaviors of speakers from filmed conversation. Because the coding method is attentionally demanding, it is difficult for non-trained coders to code specific behaviors reliably in real-speed conversation. Even for well-trained coders, there is a limitation on the number of nonverbal behaviors that can be extracted at once, which in turn causes coders typically have to watch a target film more than once in order to extract a large variety of behavior reliably. Conducting hours of coding is difficult work that requires considerable stamina. Overwork and fatigue decrease the reliability of coded data, such that frequent rest periods are required, with corresponding increases in study time and accelerating costs, depending on the number of speakers as well as the breadth of behaviors being analyzed. Using a large number of coders would address these issues by reducing the burden on each individual coder, but this unfortunately increases variance among coders. Considering that the coding method is not easy for non-trained coders, there would be no problem if there are already a lot of trained coders, otherwise using a large number of coders may result in lower inter-coder reliability. To obtain a reliable data from these non-trained coders results in a high cost of training, which just does not make any sense.

An alternate method of identifying and quantifying nonverbal behavior is behavioral ratings (Guerrero 2005) that capture subjective estimates of how frequently a target behavior occurred (e.g., 1 = *never* to 7 = *always*). For example, Guerrero (1997) used such ratings data to examine interaction differences (e.g., proximity, touch, gaze, silence) between friends and romantic partners. Although both the coding and behavioral rating methods are based on human judgment, the former extracts an actual measured value of the specific behavior on a time scale of seconds, while the latter reflect raters' subjective judgment using an evaluation scale. Compared to the coding method, which requires that focused, specific attention be paid to identify specific behavioral instances, the behavioral rating method allows attention to be allocated to the whole speaker, which permits the researcher to evaluate more aspects of nonverbal behavior simultaneously. This reduces the amount of time required for extracting behavior. But it should be noted that third-person behavioral ratings carry their own set of limitations. Rating data represent the subjective view of the rater and thus could merely reflect impressions of the behavior in question. There could of course also be individual variations in judgment. Compared with more objective measurements and counts of behavior types, subjective judgments in the form of more general ratings (e.g., *smiled a little* or *smiled a lot*) would be expected to vary more across different raters, introducing a considerable source of error. To solve this problem, the effort required to train raters cannot be ignored, and increasing the number of raters would cause the same problem as using a large number of coders. Thus, behavioral ratings do not reduce manpower requirements after all.

Researchers have addressed the problem of the time cost of extracting behaviors by presenting only brief extracts of filmed interactions and asking observers to rate the quantity

and nature of the behaviors shown; characteristics of the behaviors identified in the excerpts relative to those of the longer originals (e.g., representativeness, efficacy) are then evaluated (e.g., Ambady et al. 2002; Ambady and Rosenthal 1992, 1993). Ambady et al. (2002) explored the link between health care providers' nonverbal communication patterns and therapeutic efficacy using brief 20-s clips that were extracted from longer films depicting therapy sessions. These brief samples of therapists' nonverbal behavior at admission correlated with clients' improvements in physical and cognitive functioning at discharge and 3 months following discharge. Murphy (2005) found that brief (1-min) segments of nonverbal behaviors (gazing, gestures, nods, smiles, and self-touch) were significantly correlated with behaviors across the longer (15-min) interaction from which they were extracted. Murphy's results suggest that using "thin slices" of behaviors to represent those across a longer span of time could provide significant cost savings by reducing coding requirements.

Although the "thin slice" technique reduces coding resources, multiple coders must still be trained and employed, and behavior segments must be extracted and evaluated in relationship to the longer interaction. The present study introduces an alternative method that does not require multiple coders or the extraction of segments. An automated tool using video images and speech analysis software programs captures bodily movement and vocalizations automatically in order to provide objective measurements of nonverbal behaviors. Our automated extraction procedure employs video image analysis software, which automatically calculates saturation and lightness in a target film. Changes in these variables are used to operationally define body movements of a speaker and are thus quantified as nonverbal behaviors. A microphone is used to record sounds, and speech-signal analysis software is run to automatically assess whether vocalizations occurred. The more traditional coding and behavioral rating methods used mean scores across coders in data analysis, assuming that reasonable levels of inter-coder reliability are obtained, as assessed using correlational statistics such as α or κ coefficients (Baesler and Burgoon 1987). On the other hand, automated tools such as computer programs allow all users to obtain the same measurements from a behavior sample with minimal risk of error, assuming that all operators use the same program. Such an approach would in theory maximize data extraction accuracy while obviating the need for expensive training and effort of multiple coders.

It is crucial to demonstrate that (1) measures of nonverbal behavior derived using software programs correspond to those extracted via the more traditional method of coders/raters and (2) that software-derived behavior provides a useful predictor of other variables of interest (e.g., interpersonal impressions). The present study compared the automated method with previously studied coding and third-person perspective rating methods. Mean scores and correlation coefficients were used to compare the methods and assess similarities/dissimilarities. We were specifically interested in the extent to which the methods yielded comparable measurements. We also examined the extent to which nonverbal behaviors as extracted using the various methods predict interpersonal impressions based on both third person and conversational partner perspectives. This study had no a priori hypotheses, given that it represents the first comparison of the three extraction procedures.

Method

This study was conducted in three steps: (1) conversation session, (2) extraction of nonverbal behavior using the three methods described above, and (3) collection of impression ratings of each speaker from the third person and conversational partner points of view.

Conversation Session

Seventy Japanese undergraduates (28 females, 42 males, *Mean age* = 18.65, *SD* = 0.70) participated as speakers in exchange for extra course credit. Each speaker was asked to get acquainted with a same-sex stranger by engaging in a 6-min conversation; conversation topics were not specifically set. The dyads were seated opposite one other, 80 cm apart, and their interaction was videotaped using a camera (HDR-SR12; SONY) placed 250 cm away and to the right side of the speakers. Directional headset microphones were used to record separate utterances (ATM75; audio-technica). After the conversation, speakers provided impression ratings of their partners by completing a Japanese property-based adjective measurement questionnaire (Hayashi 1978). The scale was made to assess general impressions of persons and was based on Osgood's three dimensions of meaning (e.g., Osgood and Suci 1955) and Rosenberg's dimensions of personality impression (e.g., Rosenberg et al. 1968). The questionnaire consists of 20 pairs of adjectives (e.g., pleasant/unpleasant, prudent/imprudent, active/passive), each rated on a 7-point scale. The adjectives load onto three factors based on a formal factor analysis: familiarity, social desirability, and activeness. Speakers were then debriefed and thanked for their participation.

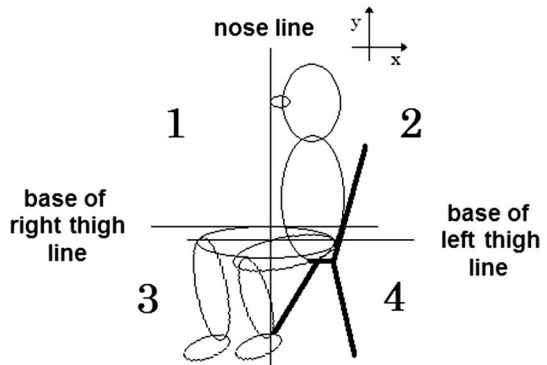
Extraction of Nonverbal Behavior

Three types of nonverbal behavior were extracted: hand gestures, adaptors, and utterances. These behaviors were chosen because they indicate interest in, as well as stress experienced during, a conversation (Ekman and Friesen 1972; Ito 1991), but also because they are discriminative and easily observed in the seated position. A hand gesture is used to illustrate or punctuate speech and may have a direct verbal translation. Previous studies suggest that hand gestures accompanying speech convey information to a listener (e.g., Krauss et al. 1991) and also indicate interest or a positive attitude during the conversation (Ito 1991). On the other hand, adaptors often involve unintentional touching of one's own body and indicate stress or negative attitudes (Ekman and Friesen 1972). In addition to these behaviors, verbal utterances were measured.

Automated Method Using Video Image and Speech-Signal Analysis Software

This study used the same paradigm as Fujiwara and Daibo (2010), digitally videotaping the speakers and employing two-dimensional motion analysis software to analyze the tape (Dipp-Motion XD Ver.3.20-2). This software can automatically track and capture two-dimensional body movements in chronological order. For each speaker, coordinate points for the fingertips, nose, and base of both thighs were first captured every 0.5 s in chronological order. Second, positions of both hands were calculated (see Fig. 1). If the fingertip of the right hand was above the base of the speaker's right thigh and in front of the nose (area 1 in Fig. 1), the movement was defined as a "hand gesture". In contrast, if the fingertip of the right hand was above the base of the speaker's right thigh and behind the nose, that is, at the speaker's side (area 2 in Fig. 1), the movement was defined as an "adaptor". If the fingertip was under the base of the speaker's thigh (areas 3 and 4 in Fig. 1), the movement was regarded not as a specific action but rather as basic positioning (e.g., a hand was on the person's lap or knee), and the movement was eliminated from further analysis. The total times spent making hand gestures and engaging in adaptors were used in subsequent analyses. In addition, frequency of utterances was computed with the aid of a specialized utility program in the TalkAnalyzer 1.2.5 software package, which can

Fig. 1 Definition of movements using the position of each hand (Fujiwara and Daibo 2010)



code speech (i.e., verbalizations) into 1–0 data points every 0.5 s in chronological order. This software uses loudness of sound input to a personal computer (with the threshold set by the user) in order to evaluate whether a verbalization occurred. If loudness is above a threshold value, the software codes “1”; otherwise, it codes “0”. Total time spent making utterances was analyzed.

Coding Method Using an Event Recorder

A “Sigsagi” event recorder (Arakawa and Suzuki 2004) was created using Visual Basic 6.0 (Microsoft). This event recorder codes a “1” if the user presses a particular key; in the absence of a key press it codes a “0” every 0.5 s in chronological order. Four coders (three undergraduates and a graduate student) recorded the occurrence of hand gestures, adaptors, and utterances. In order to reduce coder burden, speakers were distributed across coders using a ${}_4C_2$ combination to ensure that two coders extracted data from every speaker.¹ To establish reliability between coders, each coder independently engaged in extracting specific behaviors from the videotaped interactions. Operational definitions of nonverbal behaviors were based on earlier studies conducted in Japan (Kimura et al. 2010), which referred to Bernieri et al. (1996) and Tickle-Degnen and Rosenthal (1987). Cases in which a hand was resting on the lap or knee were regarded as basic positioning. A hand gesture was defined as a nonverbal act that has a direct verbal translation or is used to illustrate or punctuate speech. Adaptors involved the touching of the speaker’s own head, face, arms (include crossing the arms), or legs, and lacing the fingers together on the lap or knee. Utterances involved the production of sound, not including throat clearing or audibly breathing through the nose.

Behavioral Ratings

Two raters also subjectively assessed hand gesture, adaptor, and utterance frequencies. These raters were graduate students who did not engage in the coding described above. The speakers in each conversation were rated one at a time (the other speaker was hidden from view), with each behavior rated on a 7-point scale (1 = *little* to 7 = *much*). As in the

¹ Twelve speakers each were assigned to coders A and B, coders A and C, coders A and D, coders B and C, and coders B and D. The remaining 10 speakers were assigned to coders C and D.

coding method, two raters independently engaged in rating. Each rater assessed all of the speakers.

Third Person and Conversational Partner Subjective Ratings

In order to examine the relationship between nonverbal behavior and reactions to the speaker, interpersonal impressions of the speakers were assessed.

Conversational Partners' Ratings

Conversational partners completed questionnaires after the conversations were terminated. As described above, they rated their impressions of their partner by responding to 20 pairs of adjectives (Hayashi 1978).

Third Person's Rating

Four undergraduate observers who had not previously seen the conversation films rated friendliness, seriousness, and activeness of the individual speakers in each film, using 7-point scales (1 = *disagree* to 7 = *agree*). These three items were based on items from Hayashi's (1978) assessment of three dimensions of interpersonal perception: familiarity, social desirability, and activeness. Although three scores may be obtained by averaging across 20 adjectives, a single item was chosen from each dimension and used here, to help lighten the observer's burden. Each observer independently assessed all of the speakers and provided a behavioral rating.

Results

Comparison and Correlations of Nonverbal Scores Among the Methods

The inter-coder reliabilities using the event recorder were all higher than $r = 0.82$,² so the scores for each nonverbal behavior were averaged across coders. Similarly, inter-rater reliabilities for the behavior ratings were all higher than $r = 0.73$,³ and these were averaged across raters as well (Table 1). Because the distributions of the nonverbal data were skewed, a non-parametric analysis was appropriate.

² Reliability was calculated as the correlation between the two coders. Correlation coefficients for each pair of coders were all significant ($ps < .01$): coders A and B (hand gestures $r = 0.96$, adaptors $r = 0.87$, utterances $r = 0.93$), coders A and C (hand gestures $r = 0.88$, adaptors $r = 0.97$, utterances $r = 0.95$), coders A and D (hand gestures $r = 0.99$, adaptors $r = 0.98$, utterances $r = 0.96$), coders B and C (hand gestures $r = 0.82$, adaptors $r = 0.97$, utterances $r = 0.92$), coders B and D (hand gestures $r = 0.98$, adaptors $r = 0.94$, utterances $r = 0.83$), and coders C and D (hand gestures $r = 0.88$, adaptors $r = 0.89$, utterances $r = 0.99$).

³ As in the case of the coding method, reliability was calculated as the correlation between two raters. Correlation coefficients were all significant ($ps < .01$): hand gestures $r = 0.74$, adaptors $r = 0.78$, utterances $r = 0.73$.

Table 1 Mean scores (SD) and ranges of nonverbal behaviors extracted using different methods

Behavior	Automated method		Coding method		Behavioral rating	
	Means (SD)	Range	Means (SD)	Range	Means (SD)	Range
Hand gestures	46.19 (39.24)	0–181.50	45.01 (36.49)	0–141.50	2.88 (1.52)	1–6.5
Adaptors	44.53 (61.36)	0–345.00	71.98 (68.39)	0–334.75	4.22 (1.82)	1–7
Utterances	168.26 (43.90)	39.5–252.50	196.82 (40.93)	89.5–278.25	4.62 (1.18)	2–7

$n = 70$. The automated and coding methods measured objective time spent generating each behavior. The behavioral rating score is a subjective assessment from two raters. Measures are based on raw data

Comparison of Nonverbal Scores Between Automated and Coding Methods

Initially, objective measurements of each behavior extracted using the automated and coding methods were compared in terms of time scores. On the other hand, because the scoring of behavior ratings was made using a different metric (a 1–7 scale), comparisons of the behavior ratings to the time scores were not appropriate. A Wilcoxon signed-rank test revealed that adaptor and utterance scores differed significantly between the automated and coding extraction procedures ($V_s < 339.5$, $ps < .01$, effect sizes $r_s < 0.80$): Scores for the coding method were higher than those for the automated method. There were no significant differences for hand gestures ($V = 1,138.0$, ns).

Correlations Among Automated Method, Coding Method, and Ratings

Spearman rank correlations between each method for the same nonverbal behaviors were calculated. Correlation coefficients for each pair of methods were all significant ($ps < .01$): automated and coding methods (hand gestures $r_s = 0.78$, adaptors $r_s = 0.71$, utterances $r_s = 0.46$), and automated and behavioral rating methods (hand gestures $r_s = 0.71$, adaptors $r_s = 0.39$, utterances $r_s = 0.36$). Correlations between the coding and behavioral rating methods were also significant (hand gestures $r_s = 0.86$, adaptors $r_s = 0.62$, utterances $r_s = 0.69$). Furthermore, hand gestures were significantly correlated with both adaptors ($r_s = 0.60$) and utterances ($r_s = 0.35$) for the automated method ($ps < .01$). On the other hand, hand gestures were significantly correlated only with utterances for coding ($r_s = 0.58$) and rating ($r_s = 0.63$, both $ps < .01$) methods. For the automated method, hand gestures were more strongly correlated with adaptors. For the coding method and ratings, on the other hand, hand gestures showed a stronger correlation with utterances. There were no significant correlations between adaptors and utterances for any of the methods.

Relationships with Interpersonal Impressions

Interpersonal Impressions from the Third Person Perspective

Interpersonal impression ratings were obtained from conversational partners as well as four independent observers. Inter-rater reliabilities (Cronbach's α) of the four observers were all equal to or greater than 0.45 (friendliness $\alpha = 0.70$, seriousness $\alpha = 0.45$, activeness $\alpha = 0.83$). However, the relatively low reliability for seriousness led us to exclude these ratings from further analysis. The mean friendliness score was 4.60 ($SD = 0.94$), and the mean activeness score was 4.78 ($SD = 0.99$). The correlation between friendliness and activeness scores was $r = 0.64$ ($p < .01$).

Table 2 Influence of nonverbal behavior on interpersonal impressions of observers

	Observer-rated friendliness			Observer-rated activeness		
	Automated method	Coding method	Behavioral method	Automated method	Coding method	Behavioral method
Nonverbal behavior						
Hand gestures	0.23 [†]	0.36**	0.11	0.28*	0.40**	0.20*
Adaptors	0.03	-0.05	-0.06	-0.06	-0.13	-0.01
Utterances	0.13	0.16	0.47**	0.35**	0.42**	0.68**
Adjusted R ²	0.05 [†]	0.19	0.28**	0.25**	0.51**	0.67**

Each regression analysis was conducted separately. Partial regression coefficients were standardized

[†] $p < .10$; * $p < .05$; ** $p < .01$

Multiple linear regression analyses were conducted separately for the three method types to examine the influence of extracted nonverbal behaviors on interpersonal impressions. For all three methods, hand gestures and utterances had several marginal or significant effects on friendliness and activeness ratings (see Table 2). For the automated and coding methods, hand gestures had, marginal and significant effects on friendliness, respectively. For the behavior ratings, on the other hand, utterances had a stronger effect on friendliness. For activeness ratings, hand gestures and utterances had a significant effect, and utterances had a stronger effect than hand gestures, for all three methods. The highest determination coefficient was that for ratings. Adaptors did not affect friendliness or activeness ratings in any regression.

Interpersonal Impressions of the Conversational Partner

Mean interpersonal impression scores from conversational partners were as follows: familiarity, 5.24 ($SD = 0.74$); social desirability, 4.59 ($SD = 0.55$); and activeness, 4.46 ($SD = 0.58$). The correlations between familiarity and social desirability ratings and between familiarity and activeness ratings were significant ($r = 0.45$, $p < .01$, $r = 0.46$, $p < .01$, respectively) but the correlation between social desirability and activeness ratings was not significant ($r = 0.13$, *ns*).

Multiple linear regression analyses were conducted separately for the three types of extraction method. Nonverbal behaviors extracted via the automated method had a significant effect on conversational partner ratings, with measures obtained using the other methods having no such effects (see Table 3). Although hand gestures had a significant effect on friendliness for the coding method, the determination coefficient of regression remained at a marginal significance level. For the automated method, hand gestures had a significant positive effect, and adaptors had a significant negative effect, on familiarity rating. Utterances had a significant effect on familiarity and activeness ratings.

Discussion

The purpose of this study was to examine the validity and possible utility of a software-based, automated nonverbal behavior extraction procedure that produces measurements without the use of human coders. Nonverbal behaviors (hand gestures, adaptors, and

Table 3 Influence of nonverbal behavior on interpersonal impressions of conversational partner

	Partner-reported familiarity			Partner-reported social desirability			Partner-reported activeness		
	AM	CM	BR	AM	CM	BR	AM	CM	BR
Nonverbal behavior									
Hand gestures	0.32**	0.29*	0.26	0.27	0.27	0.25	0.03	0.10	0.11
Adaptors	-0.24*	-0.16	0.00	0.02	0.02	0.09	0.18	0.10	0.10
Utterances	0.31**	0.06	0.07	-0.22	-0.22	-0.20	0.34**	0.19	0.12
Adjusted R ²	0.25**	0.05 [†]	0.06 [†]	0.01	0.01	0.00	0.11*	0.04	0.01

Each regression analysis was conducted separately. Partial regression coefficients were standardized

AM automated method, CM coding method, BR behavioral rating method

[†] $p < .10$; * $p < .05$; ** $p < .01$

utterances) were extracted via this new method and these measures were compared to those derived using the more traditional event recorder/coder and subjective behavioral rating methods.

Adaptor and utterance scores were lower for the automated procedure than for the event recorder/coder procedure, although there was no significant difference for hand gestures. For adaptor scores, a difference in terms of what was coded as basic positioning may have accounted for this result. If each fingertip was under the base of the speaker’s thigh, basic positioning, as opposed to a more specific type of behavior, may have been coded by the automated method. In contrast, behaviors such as lacing the fingers together could have been counted as adaptors during event recorder extraction, which would be expected to increase counts of adaptors as compared to the automated method. No such difference was observed for hand gestures, given that these behaviors typically occur above the thigh and in front of the body. The cause of the method difference for utterances might relate to the experimental environment: Depending on how the microphone was mounted (i.e., distance from and angle to the speaker’s mouth), some speech sounds may not have registered and therefore may not have been coded by the speech-signal analysis software.

Measures of nonverbal behavior extracted using the different methods did correlate with one another. Correlations among the three methods were highest for hand gestures ($r_s > 0.71$). Correlations between the automated and coding methods were also relatively high for adaptors ($r_s = 0.71$) and moderate for utterances ($r_s = 0.44$). These results support the validity of the automated extraction procedure, insofar that results are fairly consistent with those obtained using previously established coding methods. Behavioral ratings correlated moderately with the automated method for adaptors and utterances ($r_s < 0.39$). On the other hand, correlations with behavioral ratings were higher for the coding method ($r_s > 0.62$) than for the automated method. This stronger correlation may have resulted because both of the coding/rating methods are based on human judgment. In other words, the subjective impression of specific nonverbal behaviors may not always have corresponded to the behavior based on objective bodily movement that was extracted by the automated method. Hand gestures were more strongly correlated with adaptors when the automated method was used and were more strongly correlated with utterances when the coding or behavioral ratings method was used. This difference may have derived from the way adaptors were extracted. In the automated method, hand gestures were defined by the state of the fingertip of the right or left hand being above the base of the

speaker's right or left thigh and in front of the nose (area 1 in Fig. 1), whereas adaptors were defined by the hand in the same position relative to the thigh but behind the nose (area 2 in Fig. 1). Based on this definition, if the speaker raised a hand to the front of the body in order to touch his or her hair, a hand gesture was extracted before an adaptor (with a hand gesture being extracted again if the hand returned to the basic position). Such movements might have served to strengthen the correlation between hand gestures and adaptors for the automated method.

Hand gestures and utterances extracted via each method explained the activeness impressions of the speaker from the perspective of a third person. Ito (1991) suggested that hand gestures are associated with activeness. With respect to utterances, Ogawa (2003a) indicated that a lack of silence is assessed as active from a third person perspective. On the other hand, lively conversation is considered an expression of intimacy or "immediacy" (Andersen 1985; Daibo 1990). Hand gestures extracted with the coding method, as well as subjective utterance ratings, also had a significant influence on impressions of friendliness. By contrast, adaptors had no significant effect on friendliness or activeness impression for any of the three methods. Compared to the straightforward expressiveness of hand gestures and utterances, the more subtle and less expressive adaptors would have been less impressive. Also, because the interactions were videotaped from the right side, some adaptors may have been seen as an aspect of the speaker's posture (e.g., crossing the arms). Nonverbal behavior extracted via the automated method yielded lower coefficients of determination for both friendliness and activeness, as compared to the coding method and behavioral ratings. In a sense, this comes as no surprise. Extraction procedures using an event recorder and behavioral ratings are both based on observations from the perspective of a third person. These data should be associated with interpersonal impressions that are also based on observations from a third person perspective, due to similarities in method. However, the automated method remains potentially valuable in terms of cost effectiveness and reduced need to train skilled coders.

In addition, nonverbal behavior extracted using speech-signal analysis software had a significant effect on the interpersonal impressions of a conversational partner. Utterances had significant positive effects on familiarity and activeness, supporting Ogawa's (2003b) view that lack of silence is associated with impressions of activeness from a conversational partner. In the context of a conversation to get acquainted with a stranger, introducing oneself and talking about recent personal events might be received positively. Hand gestures had significant positive effects, and adaptors had significant negative effects, on familiarity ratings, which appears to support previous suggestions that hand gestures reflect a positive attitude (Ito 1991) and that adaptors indicate stress or negative attitude (Ekman and Friesen 1972). On the other hand, nonverbal behaviors extracted using an event recorder and behavioral ratings showed few such effects. These results suggest the possibility that impressions from a conversational partner may not always be closely associated with such judgments from the perspective of a third person, with a potential advantage going to the automated method in terms of its ability to predict variance in the impressions of conversational partners.

Limitation and Future Perspectives

The results of this study require replication and generalization to other scenarios where nonverbal behaviors can be studied. First, the environment in which the conversations took place was quite restrictive. Speakers were seated opposite each other, with their interactions videotaped by one camera placed right beside the speakers. Hand gestures and

adaptors extracted using speakers' hand positions would therefore be largely dependent upon the sitting position. Speakers were seated in immovable chairs so that hand position could be judged on the basis of two-dimensional images, which could not be done if speakers were seated at an angle to one another or if the camera was placed elsewhere. Some ingenuity would be required to assess three-dimensional data captures through synchronizing and calibrating multiple cameras, which would enable judgments of the relative position of the speaker's hands toward his or her body. It would be fruitful to extract nonverbal behavior data on the basis of three-dimensional position coordinates.

It should also be noted that using only one rating of impressions, such as friendliness or activeness of speakers, is potentially problematic. These items were based on those used by Hayashi (1978) and can assess only very basic, general impressions. It remains to be seen whether nonverbal behaviors extracted using an automated method can predict more complex forms of interpersonal impression, such as performance in a debate about a topic or quality of a lecture. The present results should be extended to more complex, naturalistic situations.

This study examined the validity and possible utility of a new automated method for the extraction of nonverbal behaviors. Although the automated method currently requires a fairly restrictive filmed conversation environment, it requires neither multiple coders and raters nor the effort to train them, which would be expected to dramatically reduce costs. Although the software product used as the automated tool for measuring bodily movement cannot be made openly available, it is important to note that any software could be used for simply extracting a speaker's coordinate points or verbalizations. The idea of using relative reference points and loudness of sound in defining nonverbal behavior presents the opportunity to further develop automated methods for extracting nonverbal behavior that can be widely used.

References

- Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. H. (2002). Physical therapists' nonverbal communication predicts geriatric patient's health outcomes. *Psychology and Aging, 17*, 443–452.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441.
- Andersen, J. F. (1985). Nonverbal immediacy in interpersonal communication. In A. W. Siegman & S. Feldstein (Eds.), *Multichannel integrations of nonverbal behavior* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum.
- Arakawa, A., & Suzuki, N. (2004). Hand movements and feeling. *Japanese Journal of Research on Emotions, 10*, 56–64.
- Baesler, E. J., & Burgoon, J. K. (1987). Measurement and reliability of nonverbal behavior. *Journal of Nonverbal Behavior, 11*, 205–233.
- Bernieri, J. F., Gills, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology, 71*, 110–129.
- Daibo, I. (1990). Expression of intimacy in interpersonal relationships: Development and dissolution of relationships in communication. *Japanese Psychological Review, 33*, 322–352.
- Dunber, N. E., & Burgoon, J. K. (2005). Perception of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships, 22*, 207–233.
- Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication, 22*, 353–374.
- Floyd, K. (2006). *Communicating affection: Interpersonal behavior and social context*. Cambridge, UK: Cambridge University Press.

- Fujiwara, K., & Daibo, I. (2010). The function of positive affect in a communication context: Satisfaction with conversation and hand movement. *Japanese Journal of Research on Emotions*, 17, 180–188.
- Guerrero, L. K. (1997). Nonverbal involvement across interactions with same-sex friends, opposite-sex friends, and romantic partners: Consistency or change? *Journal of Social and Personal Relationships*, 14, 31–58.
- Guerrero, L. K. (2005). Observer ratings of nonverbal involvement and immediacy. In V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 221–235). Mahwah, NJ: Lawrence Erlbaum.
- Hayashi, F. (1978). The fundamental dimensions of interpersonal cognitive structure. *Nagoya University School of Education, Department of Education Psychology Bulletin*, 25, 233–247.
- Ito, T. (1991). The characteristics of unit nonverbal behaviors in face-to-face interaction. *The Japanese Journal of Experimental Social Psychology*, 31, 85–93.
- Kimura, M., Daibo, I., & Yogo, M. (2010). Investigation into the mechanism of interpersonal communication-cognition as a social skill. *Japanese Journal of Social Psychology*, 26, 13–24.
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61, 743–754.
- Murphy, N. A. (2005). Using thin slices for behavioral coding. *Journal of Nonverbal Behavior*, 29, 235–246.
- Ogawa, K. (2003a). The effects of utterance balance in a conversation on a neutral observer's impressions of the conversants and the conversational process. *The Japanese Journal of Experimental Social Psychology*, 43, 63–74.
- Ogawa, K. (2003b). The effects of balance in the amount of conversational utterance on the impressions of the partner and the conversation. *IECICE Technical Report*, 103, 63–74.
- Osgood, C. E., & Suci, G. J. (1955). Factor analysis of meaning. *Journal of Experimental Psychology*, 50, 325–338.
- Riggio, R. E., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology*, 50, 421–427.
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, 9, 283–294.
- Tickle-Degnen, L., & Rosenthal, R. (1987). Group rapport and nonverbal behavior. *Review of Personality and Social Psychology*, 9, 285–293.